TEORIA DA INFORMAÇÃO E DA CODIFICAÇÃO

Olivier Rioul

TEORIA DA INFORMAÇÃO E DA CODIFICAÇÃO

Tradução
José Carlos Magossi

EDITORA UNICAMP

EDITORA UnB

Grafia atualizada segundo o Acordo Ortográfico da Língua
Portuguesa de 1990. Em vigor no Brasil a partir de 2009.

FICHA CATALOGRÁFICA ELABORADA PELO
SISTEMA DE BIBLIOTECAS DA UNICAMP
DIRETORIA DE TRATAMENTO DA INFORMAÇÃO
Bibliotecária: Maria Lúcia Nery Dutra de Castro – CRB-8ª / 1724

R479 Rioul, Olivier
Teoria da informação e da codificação / Olivier Rioul; tradução José Carlos Magossi. – Campinas, SP: Editora da Unicamp; Brasília, DF: Editora Universidade de Brasília, 2018.

1. Shannon, Claude E. 2. Teoria da informação. 3. Comunicação – Princípios matemáticos. 4. Sistemas de transmissão de dados. 5. Compressão de dados (Computação) 6. Entropia. I. Magossi, José Carlos. II. Título

CDD - 003.54
- 001.510151
- 621.319
- 005.746
- 536.73

ISBN 978-85-268-1394-6 (Editora da Unicamp)
ISBN 978-85-230-1310-3 (Editora Universidade de Brasília)

Título original: *Théorie de l'information et du codage*

Printed in Brazil.
Foi feito o depósito legal.

Direitos reservados a

Editora da Unicamp
Rua Caio Graco Prado, 50 – Campus Unicamp
CEP 13083-892 – Campinas – SP – Brasil
Tel./Fax: (19) 3521-7718/7728
www.editoraunicamp.com.br
vendas@editora.unicamp.br

Editora Universidade de Brasília
SCS – quadra 2 – bloco C – nº 78,
edifício OK – 2º andar
CEP 70302-907 – Brasília – DF
Tel.: (61) 3035-4200
www.editora.unb.br – contatoeditora@unb.br

Sumário

Nota Prévia

Esta obra visa fornecer aos estudantes do segundo ciclo de universidades ou escolas de engenharia as noções essenciais de teoria da informação, para aplicações na codificação de fonte e de canal. Os estudantes aos quais a obra se destina devem possuir as bases da teoria das probabilidades (variáveis aleatórias), disciplina sobre a qual se apoia a teoria da informação.

Poucos livros consagrados totalmente ou em grande parte à teoria da informação têm sido publicados em língua francesa, apesar da importância desse assunto para o ensino em universidades e escolas de engenharia. Há duas exceções notáveis: o terceiro volume de *Introduction à la théorie de la communication*, de Élie Roubine, publicado em 1970, e a monografia de Gérard Battail, *Théorie de l'information: Application aux techniques de communication*, publicada em 1997.

Parece que a presente obra é provavelmente a única referência em *francês* que trata da teoria da informação a este nível de detalhe, desde a apresentação das ferramentas básicas da teoria (entropia, informação mútua) até a demonstração dos teoremas de Shannon (para a codificação de fonte e de canal). Diversas referências em inglês (ver a bibliografia no final desta obra) têm sido fontes de inspiração, entre elas o excelente *The Theory of Information and Coding*, de Robert J. McElice e o indiscutível *Elements of Information Theory*, de Thomas M. Cover e Joy A. Thomas. Algumas partes da presente obra resultam igualmente de trabalhos pessoais do autor, principalmente sobre o estabelecimento de uma ligação com informação de Fisher e sobre uma prova original da desigualdade da variância entrópica.

Esta obra nasceu de cursos dados pelo autor na Escola Nacional Superior de Telecomunicações (ENST), na Escola Nacional Superior de Técnicas Avançadas (Ensta), na Universidade Pierre e Marie Curie (Paris VI) e na Universidade de Paris-Sud XI. Sua redação evoluiu regularmente por mais de dez anos.

Eu gostaria de agradecer aqui a todas as pessoas, colegas e amigos que, de uma maneira ou de outra, tornaram possível a organização dos cursos dos quais este livro provém, ajudaram na sua elaboração, ou me encorajaram a publicá-lo; agradeço em particular à Gerard Battail, Jean-Claude Bic, Maurice Charbit, Gérard Cohen, Pierre Duhamel, Philippe Gallion, Georges Rodriguez-Guisantes e Bruno Thedrez.

Boa leitura!

Olivier Rioul

Introdução

Claude Elwood Shannon nasceu em Michigan (USA), em 1916. Jovem doutor, ele foi contratado em 1941 pelos laboratórios Bell (New Jersey) onde iniciou estudos profundos sobre os problemas de comunicação. Os resultados de vários anos de pesquisa foram publicados em 1948: "A Mathematical Theory of Communication", *Bell System Technical Journal*, vol. 27 (1948), pp. 379–423 e 623–656.

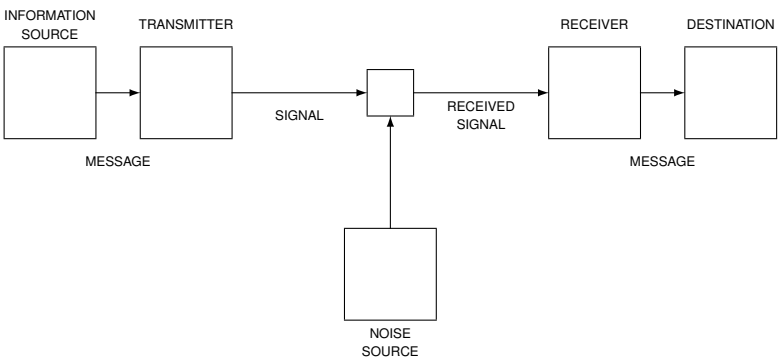

Figura 1. *O « paradigma de Shannon » (figura extraída do artigo de Shannon de 1948).*

Na introdução de seu artigo ele esquematiza um sistema geral de comunicação como na figura 1 e escreve: « *The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point* »[1].

1. « O problema fundamental da comunicação é o de reproduzir em um ponto dado, exatamente ou aproximadamente, uma mensagem selecionada em um outro ponto. »

A fim de resolver esse problema ele cria, na sequência de seu artigo, um ramo totalmente novo da matemática, atualmente chamado de *teoria da informação*. Ao fazê-lo, ele descobre não somente os conceitos fundamentais como de informação (mútua) e entropia, como prova ainda os famosos "teoremas de Shannon" que indicam os limites fundamentais, tanto para codificação de fonte como para codificação de canal.

Inicialmente, os resultados de Shannon foram tão originais que suscitaram controvérsias; alguns estudiosos da época tiveram dificuldade para compreender sua importância. Lentamente, os teoremas de Shannon foram digeridos – e foram tornando-se mais rigorosos – pela comunidade científica. Atualmente, centenas de artigos são publicados todo ano em teoria da informação. Esta obra apresenta os resultados fundamentais dessa teoria, tais como eles apareceram há mais de meio século.

Mesmo sem se basear em seu famoso artigo de 1948, Shannon é universalmente reconhecido como o fundador da teoria da informação. Ele é, mais ainda, o colaborador mais importante nessa área entre os anos de 1950 e 1960. Sua supremacia foi então enorme. Ainda hoje, numerosos são os livros ou manuais sobre essa teoria que seguem, mais ou menos, o plano do artigo inicial de Shannon:

Fonte discreta: entropia de uma variável aleatória discreta, propriedades e unicidade. entropia de uma fonte discreta (sem memória, markoviana, ergódica). Códigos de comprimento variável e teorema de Shannon ($R \geqslant H$).

Canal discreto: entropia condicional e capacidade do canal. Códigos corretores de erros e teorema de Shannon ($R \leqslant C$).

Caso contínuo: teorema de amostragem. Entropias para o caso contínuo. Capacidade (caso gaussiano: $C = W \log \frac{P+N}{N}$), taxa de distorção $R(D)$ e teoremas de Shannon associados.

Em retrospecto, tendo em conta a evolução da teoria, é possível criticar um certo número de pontos nessa apresentação.

Distinguir as ferramentas de suas aplicações

Em primeiro lugar, parece claro que Shannon foi motivado pela resolução precisa de problemas em *teoria da comunicação*, ou seja os problemas de codificação de fonte (compressão de dados) e de codificação de canal (transmissão de dados). Embora essas aplicações sejam fundamentais, podemos considerar hoje que a teoria da informação existe como um domínio próprio, com aplicações não somente em comunicação, mas também em estatística, em criptografia, em informática, em economia, em mecânica estatística, em teoria de jogos etc. (ver figura 2).

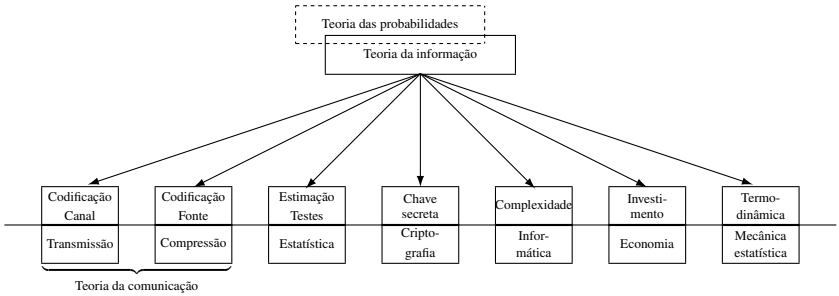

Figura 2. *Algumas conexões entre teoria da informação e outros domínios científicos.*

Dessa ótica, a *teoria da informação* pode ser vista como uma extensão da teoria das probabilidades, cujo objetivo é o estudo de ferramentas matemáticas poderosas, como medidas logarítmicas da « informação » – aplicáveis a numerosas áreas. Nesta obra, separaremos claramente a apresentação de *ferramentas* teóricas (primeira parte) de suas aplicações na *codificação* (segunda parte), como indicado na figura 3.

Unificar o discreto e o contínuo, a fonte e o canal

O plano de Shannon preconiza uma separação nítida entre o caso discreto e o contínuo. No entanto, alguns conceitos fundamentais são facilmente transpostos de um para outro. Nesta obra vamos tentar então apresentar uma visão unificada, não fazer distinção a menos que seja absolutamente necessário (propriedades da entropia, por exemplo).

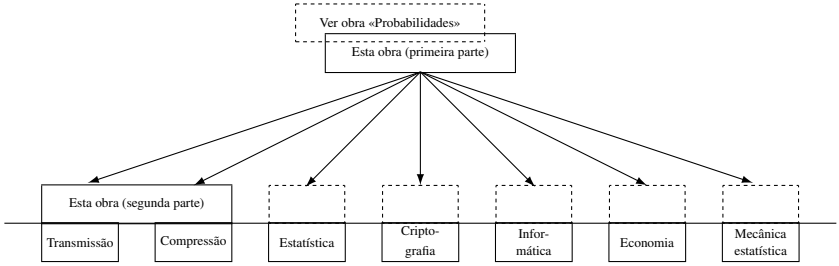

Figura 3. *Plano desta obra.*

Esse plano preconiza assim uma separação nítida entre as aplicações da teoria da informação em codificação de fonte, por um lado, e de canal por outro. Ou seja, existe uma forte dualidade entre estes dois domínios. Nós presenciamos hoje um

renascimento do interesse por aplicações de codificação conjunta fonte/canal. É por isso que adotamos aqui uma apresentação em paralelo das aplicações para codificação de fonte e de canal (em particular, as propriedades de funções taxa-distorção e capacidade-custo são mostradas simultaneamente).

Distinção entre os vários níveis teóricos e práticos

Na apresentação inicial de Shannon, encontram-se, misturadas com as ferramentas probabilísticas da teoria da informação, além de um teorema de amostragem de sinal[2], soluções práticas para codificação (códigos de comprimento variável para compressão de dados, códigos corretores de erros de transmissão). Alguns resultados importantes da teoria de codificação podem efetivamente ser demonstrados com a ajuda dessas soluções práticas: por exemplo, pode-se demonstrar o teorema de Shannon para compressão sem perdas com a ajuda dos códigos de comprimento variável (de Fano-Shannon) eficazes. Pode-se também, graças aos trabalhos de Gallager, demonstrar o teorema de Shannon para codificação de canal graças aos sutis limites sobre a probabilidade de erro de um sistema codificado. Mas as ferramentas utilizadas são então relativamente distantes daquelas originalmente previstas pela teoria da informação.

Assim, os domínios da codificação de fonte (com ou sem perdas) e da codificação de canal podem se dividir em (ao menos) três níveis, como ilustrado na figura 4. Cada nível é importante e utiliza seus próprios métodos.

1) O abordado nesta obra é teórico: se trata de introduzir os limites fundamentais (teoremas de Shannon) com a ajuda de conceitos desenvolvidos na teoria da informação.

2) Em um segundo nível, um estudo mais detalhado dos desempenhos permite não somente avaliá-los por um sistema de códigos utilizado na prática, mas também, graças aos estudos de expoentes de Gallager, reforçar as conclusões dos teoremas de Shannon.

3) Finalmente, um terceiro nível diz respeito às construções de códigos ou sistemas práticos (porque são *eficazes* no nível de codificação e decodificação). Às vezes – como para os códigos de comprimento variável – seus estudos permitem igualmente a conclusão pelos teoremas de Shannon.

Plano desta obra

A *primeira parte* apresenta as ferramentas gerais da teoria da informação.

---

2. Curiosamente, Shannon é mais conhecido, em tratamento de sinal, pelo teorema de amostragem que lhe atribuem equivocadamente; esse teorema está implícito nos trabalhos de Nyquist (1928) e de muitos outros autores. Foi demonstrado por Whittaker em 1915, e é possível até se remeter a Borel (1897).

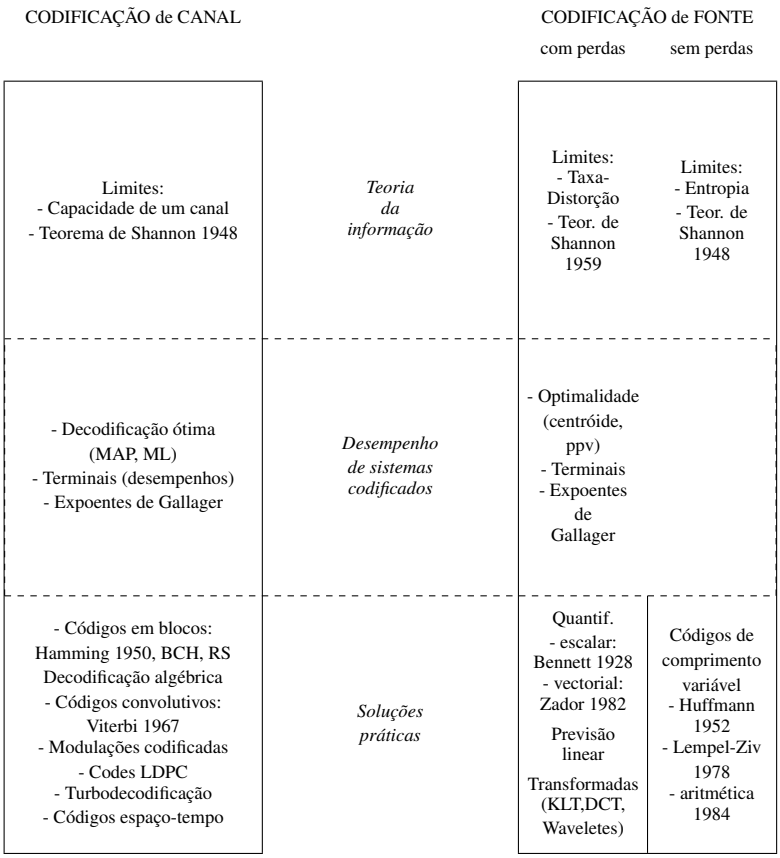

Figura 4. *Codificação de canal e codificação de fonte: organização de cursos possíveis.*

São abordadas as noções de entropia e entropia relativa (ou divergência); a propriedade de positividade da divergência é fundamental, pois ela permite, posteriormente, provar facilmente a maior parte das propriedades das ferramentas da teoria da informação. Na sequência, estuda-se informação no sentido de Shannon (informação mútua): a noção fundamental ao redor da qual toda esta obra está articulada. As diversas noções de entropia (absoluta, diferencial, condicional) são apresentadas como importantes por si mesmas e como ferramentas de cálculo que serão úteis posteriormente para aplicação na codificação. Em particular, são tratadas em detalhes as técnicas de maximização de entropia e as desigualdades de Fano. O teorema de tratamento de dados por cadeias de tratamento é uma desigualdade fundamental para sua aplicação

na codificação de fonte e canal abordada na segunda parte. Finalmente, a primeira parte termina com a apresentação de outras ferramentas (informação de Fisher, erro quadrático médio) úteis na teoria da estimação; faz-se a ligação das ferramentas da teoria da informação, *via* desigualdades da variância entrópica e identidade de Bruijn.

A *segunda parte* apresenta as aplicações da teoria da informação na codificação de fonte e canal. Primeiramente são introduzidos os modelos fundamentais de fonte de informação e de canal de transmissão, assim como as problemáticas duais de fonte e canal. Abordam-se em seguida os limites de Shannon sobre o desempenho de sistemas codificados, dados pela função taxa-distorção (para codificação da fonte) e pela função capacidade-custo (para codificação de canal). Trata-se em detalhes das propriedades gerais dessas funções e de seus cálculos para fontes e canais particulares. Finalmente, a segunda parte termina com a demonstração dos teoremas fundamentais de Shannon para codificação de fonte e de canal com a ajuda da noção de sequência típica (que utiliza a lei fraca dos grandes números).

Os *exercícios* e *problemas* completam o texto. Eles fornecem um certo número de resultados mas não são corrigidos. A resolução e a redação das soluções constituem uma parte essencial do trabalho pessoal necessário para assimilar o conteúdo desta obra. Os enunciados dos *exercícios* estão agrupados no anexo A e estão voltados para as ferramentas da teoria da informação. Os *problemas* estão agrupados no anexo B e discutem essencialmente os limites de Shannon para codificação de fonte e de canal.

PRIMEIRA PARTE

Ferramentas da teoria da informação

Capítulo 1

Entropia e entropia relativa

A teoria da informação é baseada em uma descrição probabilística de dados e de sistemas, os quais são modelados com a ajuda de *variáveis aleatórias* (forma abreviada: v.a.). Após alguns lembretes sobre variáveis aleatórias discretas e contínuas, este capítulo preliminar introduz a noção de *entropia* de uma variável aleatória, antes de expor a propriedade de *positividade da entropia relativa* (também chamada de *divergência*). Essa propriedade é fundamental, pois permite uma introdução rápida de importantes conceitos nos capítulos que se seguem.

1.1. Lembretes sobre variáveis aleatórias

Uma v.a. X é definida por *distribuição de probabilidade* $p(x)$. Nesta obra, adota-se sempre que possível a notação simplificada $p(x)$ em vez de $p_X(x)$. Assim, ao considerar duas distribuições $p(x)$ e $p(y)$, suas duas funções « p » não são as mesmas; elas correspondem respectivamente às distribuições de probabilidade p_X de X e p_Y de Y.

Há essencialmente dois tipos de variáveis aleatórias: as v.a. *discretas* e as v.a. *contínuas*.

1.1.1. *Variáveis aleatórias discretas*

Um *símbolo M-ário* é modelado por uma variável aleatória X que pode assumir M valores em uma alfabeto $\mathcal{X}$ de tamanho M. Por exemplo, pode-se tomar:

$$\mathcal{X} = \{0, 1, ..., M-1\}.$$

Acontece muitas vezes de M ser uma potência de dois; nesse caso, cada símbolo x do alfabeto $\mathcal{X}$ pode ser representado como um bloco de $\log_2 M$ bits. Por exemplo, para $M = 8$, pode-se notar:

$$\mathcal{X} = \{000, 001, 010, 011, 100, 101, 110, 111\},$$

onde cada símbolo comporta três bits. Mais geralmente se diz que $\log_2 M$ representa *o número (médio) de bits por símbolo*.

A cada símbolo x do alfabeto $\mathcal{X}$ associa-se uma *probabilidade*:

$$p(x) = \mathcal{P}rob\{X = x\},$$

compreendida entre 0 e 1, que fornece a porcentagem de chances de que X assuma o valor x. O conjunto de probabilidades $\{p(x)\}_{x\in\mathcal{X}}$ é a *distribuição* (ou *lei*) *de probabilidade* de X. Toda distribuição $p(x)$, tal que:

$$p(x) \geqslant 0 \qquad \text{e} \qquad \sum_{x\in\mathcal{X}} p(x) = 1$$

define uma v.a. discreta X sobre $\mathcal{X}$. A probabilidade de um evento $\mathcal{A}$ se calcula pela fórmula:

$$\mathcal{P}rob\{X \in \mathcal{A}\} = \sum_{x\in\mathcal{A}} p(x).$$

Nota-se que os valores de x podem muito bem serem *vetorais*: $x = (x_1, x_2, \ldots, x_n)$. Isso não muda o formalismo acima, à condição de convir que as somas do tipo $\sum_x p(x)$ sejam de fato somatórios múltiplos (n-uplas), isto é do tipo

$$\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} p(x_1, x_2, \ldots, x_n).$$

Símbolos equiprováveis

Os símbolos de um alfabeto M-ário são *equiprováveis* se eles tem a mesma probabilidade, isto é:

$$p(x) = \frac{1}{M} \qquad (\forall x \in \mathcal{X}).$$

Dito de outra forma, a variável X é *uniforme*. Intuitivamente ela corresponde ao caso mais « imprevisível »: nenhum símbolo tem mais chances de se realizar do que outro.

Símbolos binários

O caso particular $M = 2$ corresponde a um símbolo binário, também chamado *bit* – abreviação de *binary digit* (dígito binário) – que toma seus valores no alfabeto $\{0, 1\}$.